CLAIMS

1. A computer-implemented method, comprising:

receiving, at a data service engine, a query from a user, wherein the query comprises a string of characters;

identifying, by the data service engine, a plurality of candidate topics based on the query;

determining a similarity score between the query and each of the plurality of candidate topics based on a Vector Propagation On a Click Graph (VPCG) model trained based on user click data, wherein the VPCG model is trained by:

identifying a plurality of queries from a plurality of users in the user click data;

identifying a plurality of user selections in response to the plurality of queries from the plurality of users in the user click data;

determining a vector representation of each of the plurality of queries in terms of a plurality of word elements in a vocabulary, wherein the determining the vector representation of each of the plurality of queries in terms of the plurality of word elements in the vocabulary comprises representing the each of the plurality of queries based on a vector representation of each of plurality of candidate topics, a total number of user clicks or selections of a candidate topic for a query in the user click data and a total number of topics in the user click data; and

determining a vector representation of each of the plurality of candidate topics in terms of the plurality of word elements in the vocabulary, wherein the determining the vector representation of each of the plurality of candidate topics in terms of the plurality of word elements in the vocabulary comprises representing the each of the plurality of candidate topics based on a vector representation of each of the plurality of queries, a total number of user clicks or selections of a candidate topic for a query in the user click data and a total number of user queries in the user click data; and

ranking the plurality of candidate topics based on the similarity scores;

selecting one or more topics from the ranked candidate topics; and

outputting the topics via a user interface (UI).

2. The computer-implemented method of claim 1, further comprising:

before determining a similarity score between the query and each candidate topic based on a Vector Propagation On a Click Graph (VPCG) model trained based on user click data, performing a first ranking to select a second one or more topics from the plurality of candidate topics;

in response to determining that each of the second one or more topics does not meet a specified criterion, performing a second ranking based on the VPCG model trained based on user click data, wherein performing the second ranking comprises:

the determining a similarity score between the query and each candidate topic based on a VPCG model trained based on user click data; and

the ranking the plurality of candidate topics based on the similarity scores.

3. The computer-implemented method of claim 1, wherein the user click data comprise a plurality of user selections in response to a plurality of queries from a plurality of users, wherein each of the plurality of user selections comprises a respective user's selection of a relevant topic to the user's query among a plurality of candidate topics provided to the user based on the user's query.

4. The computer-implemented method of claim 1, further comprising training the VPCG model based on the user click data.

5. The computer-implemented method of claim 1, wherein the determining a similarity score between the query and each of the plurality of candidate topics based on a VPCG model trained based on user click data comprises:

determining a vector representation of the query from the user in terms of a plurality of word elements in a vocabulary;

determining a vector representation of each of the plurality of candidate topics in terms of the plurality of word elements in the vocabulary; and

calculating the similarity score between a vector representation of the query and the vector representation of each of the plurality of candidate topics.

6. The computer-implemented method of claim 5, wherein the determining a vector representation of the query from the user in terms of the plurality of word elements in the vocabulary comprising representing the query from the user based on a set of ngrams trained based on the user click data.

7. The computer-implemented method of claim 6, wherein the set of ngrams comprises a vector representation of each ngram in the set of ngrams in terms of the plurality of word elements in the vocabulary based on the vector representation of each of a plurality of queries in the user click data.

8. The computer-implemented method of claim 6, wherein the representing the query from the user based on a set of ngrams trained based on the user click data comprises representing the query from the user as a weighted sum the set of ngrams based on a vector representation of each ngram in the set of ngrams, and a weight corresponding to the vector representation of each ngram.

9. A computer-implemented system comprising:

one or more computers; and

one or more computer memory devices interoperably coupled with the one or more computers and having tangible, non-transitory, machine-readable media storing instructions, that when executed by the one or more computers, perform operations comprising:

receiving, at a data service engine, a query from a user, wherein the query comprises a string of characters;

identifying, by the data service engine, a plurality of candidate topics based on the query;

determining a similarity score between the query and each of the plurality of candidate topics based on a Vector Propagation On a Click Graph (VPCG) model trained based on user click data, wherein the VPCG model is trained by:

identifying a plurality of queries from a plurality of users in the user click data;

identifying a plurality of user selections in response to the plurality of queries from the plurality of users in the user click data;

determining a vector representation of each of the plurality of queries in terms of a plurality of word elements in a vocabulary, wherein the determining the vector representation of each of the plurality of queries in terms of the plurality of word elements in the vocabulary comprises representing the each of the plurality of queries based on a vector representation of each of plurality of candidate topics, a total number of user clicks or selections of a candidate topic for a query in the user click data and a total number of topics in the user click data; and

determining a vector representation of each of the plurality of candidate topics in terms of the plurality of word elements in the vocabulary, wherein the determining the vector representation of each of the plurality of candidate topics in terms of the plurality of word elements in the vocabulary comprises representing the each of the plurality of candidate topics based on a vector representation of each of the plurality of queries, a total number of user clicks or selections of a candidate topic for a query in the user click data and a total number of user queries in the user click data; and

ranking the plurality of candidate topics based on the similarity scores;

selecting one or more topics from the ranked candidate topics; and

outputting the topics via a user interface (UI).

10. The computer-implemented system of claim 9,the operations further comprising:

before determining a similarity score between the query and each candidate topic based on a Vector Propagation On a Click Graph (VPCG) model trained based on user click data, performing a first ranking to select a second one or more topics from the plurality of candidate topics; and

in response to determining that each of the second one or more topics does not meet a specified criterion, performing a second ranking based on the VPCG model trained based on user click data, wherein performing the second ranking comprises:

the determining a similarity score between the query and each candidate topic based on a VPCG model trained based on user click data; and

the ranking the plurality of candidate topics based on the similarity scores.

11. The computer-implemented system of claim 9, wherein the user click data comprise a plurality of user selections in response to a plurality of queries from a plurality of users, wherein each of the plurality of user selections comprises a respective user's selection of a relevant topic to the user's query among a plurality of candidate topics provided to the user based on the user's query.

12. The computer-implemented system of claim 9, further comprising training the VPCG model based on the user click data.

13. The computer-implemented system of claim 9, wherein the determining a similarity score between the query and each of the plurality of candidate topics based on a VPCG model trained based on user click data comprises:

determining a vector representation of the query from the user in terms of a plurality of word elements in a vocabulary;

determining a vector representation of each of the plurality of candidate topics in terms of the plurality of word elements in the vocabulary; and

calculating the similarity score between a vector representation of the query and the vector representation of each of the plurality of candidate topics.

14. A non-transitory, computer-readable medium storing one or more instructions executable by a computer-implemented system to perform operations comprising:

determining a similarity score between the query and each of the plurality of candidate topics based on a Vector Propagation On a Click Graph (VPCG) model trained based on user click data, wherein the VPCG model is trained by:

identifying a plurality of queries from a plurality of users in the user click data;

identifying a plurality of user selections in response to the plurality of queries from the plurality of users in the user click data;

determining a vector representation of each of the plurality of queries in terms of a plurality of word elements in a vocabulary, wherein the determining the vector representation of each of the plurality of queries in terms of the plurality of word elements in the vocabulary comprises representing the each of the plurality of queries based on a vector representation of each of plurality of candidate topics, a total number of user clicks or selections of a candidate topic for a query in the user click data and a total number of topics in the user click data; and

determining a vector representation of each of the plurality of candidate topics in terms of the plurality of word elements in the vocabulary, wherein the determining the vector representation of each of the plurality of candidate topics in terms of the plurality of word elements in the vocabulary comprises representing the each of the plurality of candidate topics based on a vector representation of each of the plurality of queries, a total number of user clicks or selections of a candidate topic for a query in the user click data and a total number of user queries in the user click data; and

ranking the plurality of candidate topics based on the similarity scores;

selecting one or more topics from the ranked candidate topics; and

outputting the topics via a user interface (UI).

15. The non-transitory, computer-readable medium of claim 14, the operations further comprising:

before determining a similarity score between the query and each candidate topic based on a Vector Propagation On a Click Graph (VPCG) model trained based on user click data, performing a first ranking to select a second one or more topics from the plurality of candidate topics;

in response to determining that each of the second one or more topics does not meet a specified criterion, performing a second ranking based on the VPCG model trained based on user click data, wherein performing the second ranking comprises:

the determining a similarity score between the query and each candidate topic based on a VPCG model trained based on user click data; and

the ranking the plurality of candidate topics based on the similarity scores.

16. The non-transitory, computer-readable medium of claim 14, wherein the user click data comprise a plurality of user selections in response to a plurality of queries from a plurality of users, wherein each of the plurality of user selections comprises a respective user's selection of a relevant topic to the user's query among a plurality of candidate topics provided to the user based on the user's query.

17. The non-transitory, computer-readable medium of claim 14, wherein the determining a similarity score between the query and each of the plurality of candidate topics based on a VPCG model trained based on user click data comprises:

determining a vector representation of the query from the user in terms of a plurality of word elements in a vocabulary;

determining a vector representation of each of the plurality of candidate topics in terms of the plurality of word elements in the vocabulary; and

calculating the similarity score between a vector representation of the query and the vector representation of each of the plurality of candidate topics.